More Than a Kid's Toy: Using NeRFs for 3D Mesh Creation

Graham Webster, Hunter Stinson
Integration Innovation Inc. (i3)
Huntsville, AL
graham.webster@i3-corps.com, thomas.stinson@i3-corps.com

ABSTRACT

The paper delves into a comparative analysis of Neural Radiance Fields (NeRF) and photogrammetry from a 3D Artist's perspective, exploring the two methods' strengths and weaknesses. Initially, a brief discussion will trace the historical trajectories of both techniques, highlighting their evolution and contributions to the field of 3D scene reconstruction. It is important to note that NeRFs do not generate 3D in the same way as photogrammetry traditionally does, and we will be emphasizing NeRF's capability to synthesize realistic views from sparse input data vs. photogrammetry which can handle hundreds and even thousands of pictures. Visual examples contrasting NeRF and photogrammetry Mesh Exports are provided to compare the differences in their output representations. The potential synergy between NeRF and photogrammetry is explored. We explore avenues for NeRF use cases in 3D capture, simulation, extended reality, and real-time scene visualization applications. Technical considerations and limitations inherent in NeRF generation are discussed, including the challenges of achieving high-quality reconstructions and the suitability of NeRF for various applications. The automation of NeRF generation and the development of a user-friendly web-based viewer are briefly outlined, along with considerations and insights into the tools and techniques utilized in NeRF generation research. Finally, the paper describes potential future use cases in NeRF research, identifying areas for improvement and innovation in NeRF 3D generation methods and applications. It underscores the importance of addressing technical hurdles and maximizing the potential of NeRF for advancing 3D scene reconstruction and visualization technologies. It loosely touches on where NeRFs and future alternatives may go, including Gaussian Splatting, which has some differences from traditional NeRFs. The paper also briefly explores alternatives to NeRF, such as Gaussian Splatting, highlighting their differences from traditional NeRFs and their potential contributions to the field.

ABOUT THE AUTHORS

Graham Webster is a Full Sail University alumni who works as a Senior 3D Artist at Integration Innovation Inc. and is an avid Extended Reality (XR) designer. He has worked on numerous 3D modeling projects involving UH-60 Blackhawk variants, the MQ-9 Reaper, and land-based vehicles such as Strykers. He is responsible for asset creation, 3D optimization, 3D scene creation, effects, shaders, and lighting. He has experience with rapid prototyping and User Interface design in Augmented Reality and Virtual Reality using devices such as HoloLens 2, Magic Leap 2, and several VR headsets. Aside from his general focus on 3D and Extended Reality, he also specializes in researching new techniques and technology that benefit the military training and simulation space.

Hunter Stinson is a software engineer with experience in modeling and simulation, uncrewed systems, and Extended Reality devices. He is currently a Senior Technologist at Integration Innovation Inc. in Huntsville, AL where he helps drive technical solutions in the training and simulation domains. Hunter earned a B.S. in Computer Engineering from the University of Alabama in Huntsville and an M.S. in Software Engineering from the University of Alabama in Huntsville.

INTRODUCTION

Mesh Models and CAD models for Real-Time Applications

The use of 3D models is commonplace for immersive training and simulations in the Department of Defense. While incredibly valuable, hands-on training with a system or task trainer is limited due to the cost or availability of the system. Task training in an immersive, all-digital environment using game engines and 3D models is a means to reduce the needed hands-on time and increase training availability. However, creating 3D models is time-consuming and costly, especially when done by hand. Techniques like photogrammetry have been used to create models more quickly, and they rely on stitching photographs together to generate the model. Recently, AI-based methods like Neural Radiance Fields (NeRF) have started to provide alternatives, potentially offering improvements over traditional 3D modeling methods. Despite the promise, these techniques come with their own limitations. This paper assesses NeRF and photogrammetry for 3D model creation, specifically in the context of the Department of Defense, which uses game engine technology to enhance the realism and effectiveness of immersive training environments.

Many are familiar with CAD modeling and its role in industrial manufacturing. However, mesh 3D models are a staple of the military simulation, film, and gaming industries that now heavily rely on 3D content for their productions. For several reasons, CAD is not utilized directly by the previously mentioned industries. One primary reason is that game engine related pipelines require models optimized for real-time rendering, which often involves simplifying geometry, reducing polygon counts, and optimizing textures and materials to achieve acceptable performance while maintaining visual quality. CAD models are not usually optimized for these requirements out of the box. CAD models are typically represented as arcs, and splines as BREP data (Boundary Representative). Mesh models are more of a cage of topology, a wireframe that creates the object's shape **(See Figure 1)**. The denser the wireframe, the higher the polycount, and the less performant the asset will be at runtime. Optimization issues resulting from photogrammetry and NeRF are also an issue. CAD, photogrammetry, and NeRF produce 3D models with too many polygons to be performant in real time. Another reason CAD is not directly utilized for real time simulations is that it cannot be rigged for animation, whereas clean optimized mesh models can. Where do NeRF and photogrammetry come in handy for 3D artists if the result is rarely riggable or performant? 3D modelers can still use the result of the scan data to model over to ensure their work is proportionate and accurate, which is critical in the military training and simulation fields. It is also possible for CAD modelers to use photogrammetry and NeRFs as reference for their work, such as those in fields related to reverse engineering. With photogrammetry data and NeRF, we compare factors such as mesh density, mesh accuracy, precision, and total form. Ideally, a quality mesh model for real-time applications should match the object's form, unwrapped topology for more accurate material, and have optimized topology that does not degrade the form of the object, accuracy, scale, and proportions, which can be gained from good scan data.
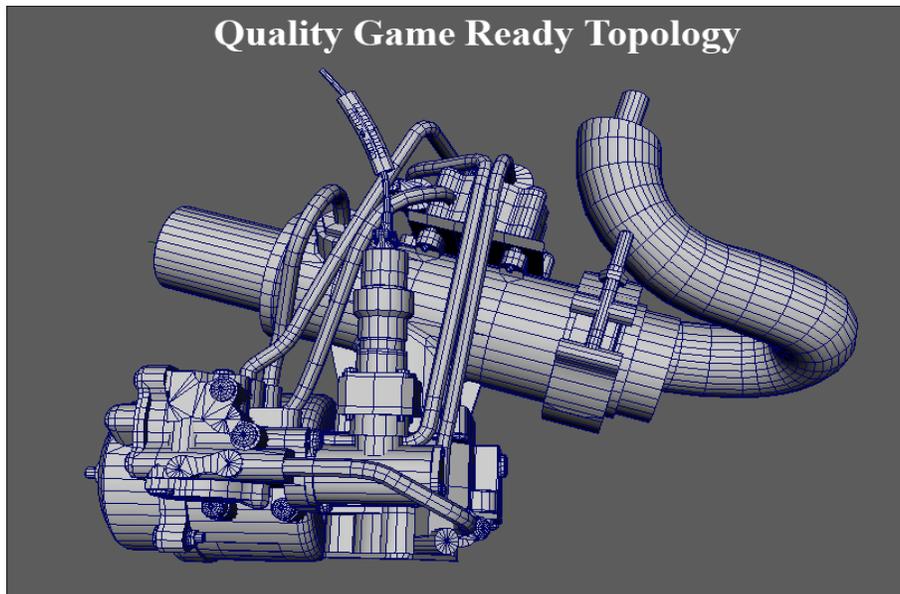

Figure 1: Game Cage Topology Example

Neural Radiance Fields

In 2020, NeRFs started gaining popularity as an upcoming method of 3D capture. The writers of the initial researchers described NeRFs as "a novel method for synthesizing novel views of complex scenes from a sparse set of input images." (Mildenhall, 2020), but what does that mean? Before the invention of NeRFs, photogrammetry was one of the major methods of 3D capture, blue light, and LiDAR scanning. NeRFs are a method that utilizes deep learning for 3D scene creation and 3D object capture. NeRFs work by representing a scene as a Radiance Field that represents scene geometry, lighting conditions, and reflectance of objects. Captured with image sets similar to photogrammetry, the NeRF typically needs to be retrained for different 3D objects or environments. Each image in a data set provides camera position/orientation data that feeds into camera ray marching to generate a set of 3D points, each with a given radiance direction. NeRFs allow reflection and shadow variation that realistically varies between positions around the real-world object by encoding individual points with multiple values.

Photogrammetry

Photogrammetry has quite a few fundamental differences at a general level compared to NeRFs. Digital photogrammetry has been a mainstay in proportional 3D capture for decades, but its concept has existed for centuries. French physicist Dominique Francois Arago is widely regarded as one of the first people to use photogrammetry with a machine called the daguerreotype (Saif, 2022). photogrammetry in the digital realm uses camera information such as camera position and orientation and the overlap of sequential photos to determine the dimensions and positions of objects. Photogrammetry will bake material information into a RGB texture map, often called a diffuse texture map. This prevents photogrammetry from accurately representing material qualities such as reflection changes because it bakes highlights and shadows into a single static map that cannot change dynamically. There are many differences between NeRFs and photogrammetry. However, a significant difference is that NeRFs use deep learning to determine 3D scenes and 3D objects without feature matching. In contrast, photogrammetry reconstructs models by triangulating shared points across multiple photographs in a data set.

NERF TOOLS AND LIMITATIONS

Choosing an application to process NeRFs is an essential first step to generating a 3D mesh. A few applications include Luma.AI, InstantNGP (also called instant NeRF), and NeRF Studio. For our research, we chose InstantNGP due to its ease of setup and the ability to view and process with on-premises computers, allowing us to test data sets more securely (Müeller, Evans, Schied, & Keller, 2022). In addition, previous research comparing InstantNGP to common photogrammetry software packages provided promising results in building 3D models under weak geometry conditions (Pepe, Alfio, & Costantino, 2023). We did not select Luma.ai because we obtained better results with an actual 4k fixed aperture camera than with video on an iPhone or images from an iPhone. Luma.ai is cloud-based, and while powerful, it is not approved to handle sensitive data sets commonly used in the defense industry. A significant factor in why we did not choose NeRF Studio despite being fully open source was that it was the most complicated of all options to set up. NeRFs have a limitation regarding the amount of images that can be processed, something that photogrammetry does not share. Too many images in a NeRF data set will cause performance issues within the NeRF viewer.

When processing a NeRF data set, we observed that 50 to 150 photos is an appropriate range of images that balances quality and processing time. It is possible to process over 150 photos if sufficient hardware resources, such as a high-end graphics card, are used. For example, we were able to process 200 images using an Nvidia RTX 4090 before noticing any performance issues. With NeRFs, we found that the image sets that worked the best were done on parts that had a logical order to the capture process. Where the photographer took photos about every 10 degrees around the subject and did a logical pass from top, middle to bottom or vice versa. NeRFs, in our testing, did not work well with a data set containing multiple subject poses. With photogrammetry, a subject can be posed, say in a pose that shows the underside of a subject, and a pose that shows the rest of it, and those poses can be correlated in software such as Reality Capture using the same data set. NeRF does not seem to operate in this manner. Having multiple poses of the same subject in NeRF viewers will cause odd results, combining both poses as if they were one object.

EXPERIMENT SETUP

The following hardware and software were used to capture, process, and evaluate NeRF and photogrammetry data.

Computer Hardware and Software

A computer with a high-end graphics card was used to process both NeRF and photogrammetry inputs. For example, we observed reduced processing time when switching to a computer with a NVidia RTX 4090 graphics card compared to an RTX 3070 Ti. Table 1 contains the hardware and software used during the experiments.

Table 1: Computer Hardware and Software Specs Used in the Experiments

Hardware Specs		Software	
Operating System	Windows 10 64-bit 10.0.19045	InstantNGP Jan 8,2023 release	Used to process NeRF images.
Graphics Card	Nvidia RTX 4090	Python 3.7+	Required to execute data transformation scripts for use with InstantNGP.
CPU	Intel Core i9-13900KS	COLMAP	Required to transform captured images and video into a format that InstantNGP understands.
RAM	64GB	Luma.ai	Used for initial experimentation, but swapped to InstantNGP due to phone resolution limitations.
Hard Drive	2TB NVMe Solid State Drive	Reality Capture 1.4	Used for photogrammetry image processing.

Capture Devices

Nikon D5600 and Nikon D3400 DSLR cameras were used to capture images. The camera had a fixed aperture to reduce lens warping. Lens warping can cause artifacts to show up in the result of a NeRF or photogrammetry capture. A 4k image resolution ensured that sufficiently detailed images were fed into the experiments.

VISUALLY COMPARING PHOTOGRAMMETRY & NERFS

Despite photogrammetry's longer history and maturity as a technology, NeRFs offer distinct advantages that traditional photogrammetry does not support. However, photogrammetry also boasts significant advantages that NeRFs have yet to address fully. By showcasing visual examples of NeRFs and photogrammetry outputs using datasets from tools like InstantNGP and Reality Capture, we can illustrate the comparative results and highlight the strengths and weaknesses of each approach. This comparative analysis aims to provide insights into the evolving landscape of 3D scene reconstruction technologies and their respective applications.

It was mentioned in the introduction what creates a quality 3D mesh model, and we loosely talked about precision and accuracy being critical for photogrammetry and NeRFs. No agreed way to quantify a photogrammetry scan's quality level and error tolerances exists, but new research is emerging to help rectify this situation (di Filippo, Antinozzi, Cappetti, & Villecco, 2023). Our research focuses on providing a comparative analysis between NeRF and traditional photogrammetry using visually observable metrics. Due to the nature of visual observation, some subjectivity exists in the evaluation results. An essential evaluation criterion is the accuracy and detail level of the 3D model generated from photogrammetry and NeRF. This means we look at what areas are entirely omitted or holes created due to shadows or highlights in the photogrammetry or over-complex reflections in NeRF. This is important as a CAD modeler or 3D artist will have an easier time recreating the scan data if the form is fully comprehensive without missing too many areas of the object. Another factor used in the comparisons is the overall mesh precision and accuracy of photogrammetry vs NeRFs. If the form of a scan subject is comprehensive and the mesh is both dense and accurate, then that is ideal for an engineer or 3D modeler to work from. Another factor we look at is how well the materials and lighting are represented and how that may impact the quality of the mesh. For a simulation 3D artist,

the quality of textures and lighting is essential when creating realistic textures, and accurately capturing the original lighting conditions of a scene will help create well-lit and realistic scenes.

MATERIALS AND LIGHTING

The following figures are examples of the material quality variation between NeRF and photogrammetry results.
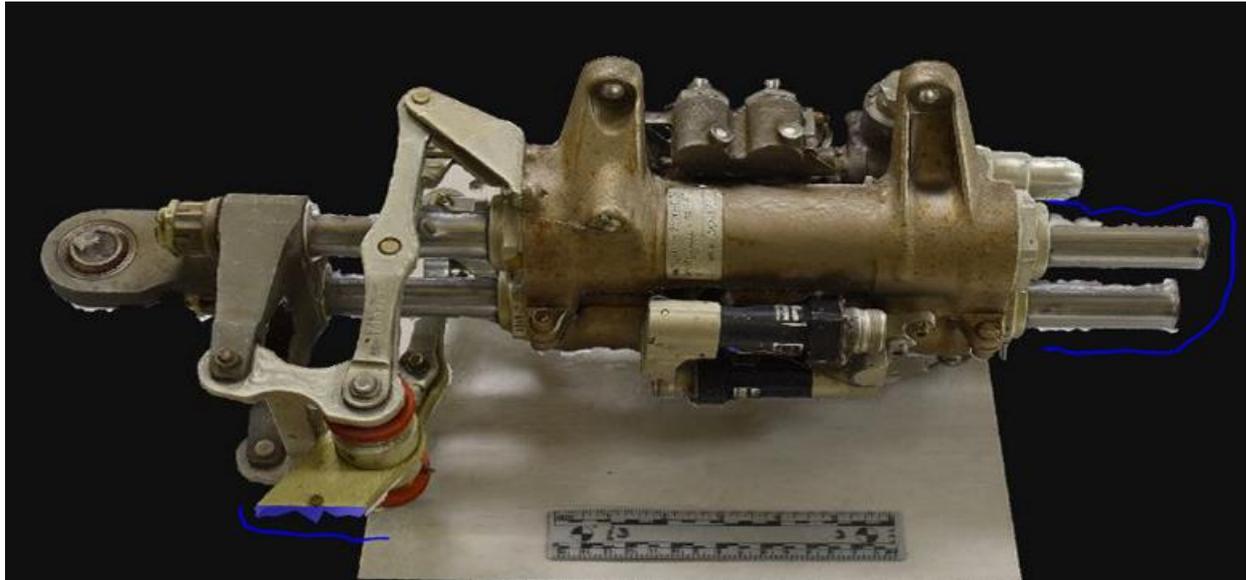

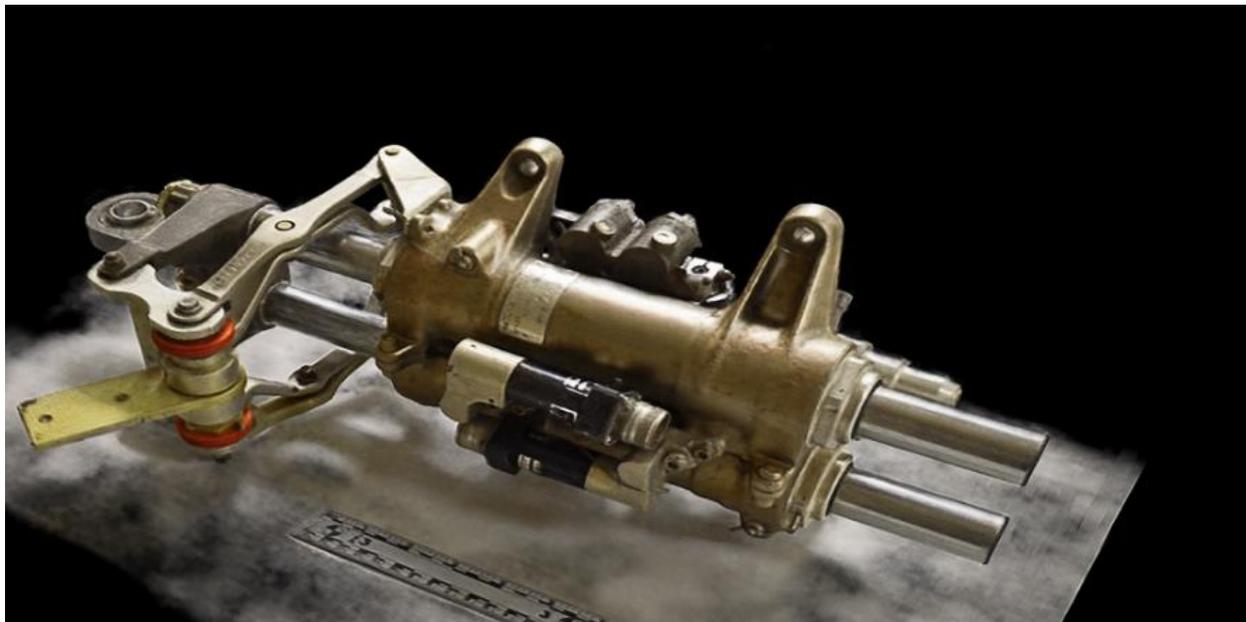**Figure 2: Photogrammetry Materials & Lighting**

Figure 3: NeRF Materials & Lighting

NeRF Pros: Materials & Lighting

At first glance, the photogrammetry example **(See Figure 2)** appears relatively muted in its reflection values. Since these examples are static images, it is important to note that the reflections, and shadows, are static to the model in the photogrammetry example, and the shadows are static as well. The photogrammetry software is baking in color, shadows, and reflections into one texture map, referred to as a diffuse map. Typically speaking, reflections and shadows are handled as separate textures so they can render the correct passes needed to represent those material qualities realistically. With the NeRF example **(See Figure 3)**, it becomes visually apparent that reflections on metallic and highly reflective surfaces are being represented in a much more realistic looking way. If one were to rotate this model around in the NeRF viewer, they would see the reflections accurately updated based on the camera position. This visually shows one of the major benefits of NeRFs: they better handle different material reflectivity. NeRF's camera view dependency can handle simple reflections (Yuan-Chen Guo, 2021). This shows up well when looking at curved reflective surfaces. This is useful for a few things such as allowing artists to better understand the lighting conditions around their subject for more realistic visualization in a simulation. It is worth noting that NeRF software can still bake textures into a diffuse map for artists to apply in modeling software if they desire.

NeRF Cons: Materials & Lighting

One downside to NeRF is that it does a decent job of simple highlight reflections, but confusing results can occur when a flat mirror-like object is in the camera shot. Flat mirror-like surfaces with complicated reflection calculations may cause confusing results (Yuan-Chen Guo, 2021). Flat mirror-like surfaces often have inaccurate depth estimations between views. The object depicted below (See Figure 4) is supposed to be a flat piece of metal with a skull cut out, but the depth is not planar in the NeRF result.

Figure 4: NeRF Planar Reflection Issues

Another downside to NeRF materials and reflections that photogrammetry shares is the inability to properly capture transparent objects **(See Figure 5)** such as glass or see-through plastic. The only factor that allowed the windshield in Figure 4 to be captured is that it is tinted enough to have reduced transparency, but even the edges came out extremely rough. Suppose a transparent object needs to be captured with NeRF or photogrammetry. In that case, it is best to

utilize evaporating developer spray to mute the reflections and transparency so that the software can better understand the form of the subject.

Figure 5: NeRF with Tinted Glass

MESH FORM AND MESH PRECISION

When the data sets for photogrammetry and NeRFs are captured with proper camera equipment, and the photographer considers the dependencies for both one typically sees better mesh precision from photogrammetry. Issues with the total form of mesh models created from photogrammetry occur with reflective surfaces in the results once metallic surfaces get more reflective. The mesh model exports without materials and textures (See Figure 6) shows that the mesh precision is higher in the top example. One can see this by simply noting that the photogrammetry mesh has less noise on the mesh. Visual examples aside, it is also worth noting that the precision of camera-based photogrammetry has already been assessed in several research papers. They determined the precision and accuracy of reconstruction by comparing them to the ground truth measurements of real-world objects to photogrammetry results. They found sub-millimeter to millimeter level accuracy of scanned objects (Croce, et al., 2024). The NeRF mesh model depicted in **Figure 6** shows that the highly reflective parts were not entirely omitted. Looking at the blue circled areas in the photogrammetry example, the photogrammetry did not capture the full form of the subject due to the method's limitations with reflective metallic surfaces. This highlights something worth noting: NeRFs do an excellent job at capturing the full form of an object or scene, despite the precision not being on par with photogrammetry yet.
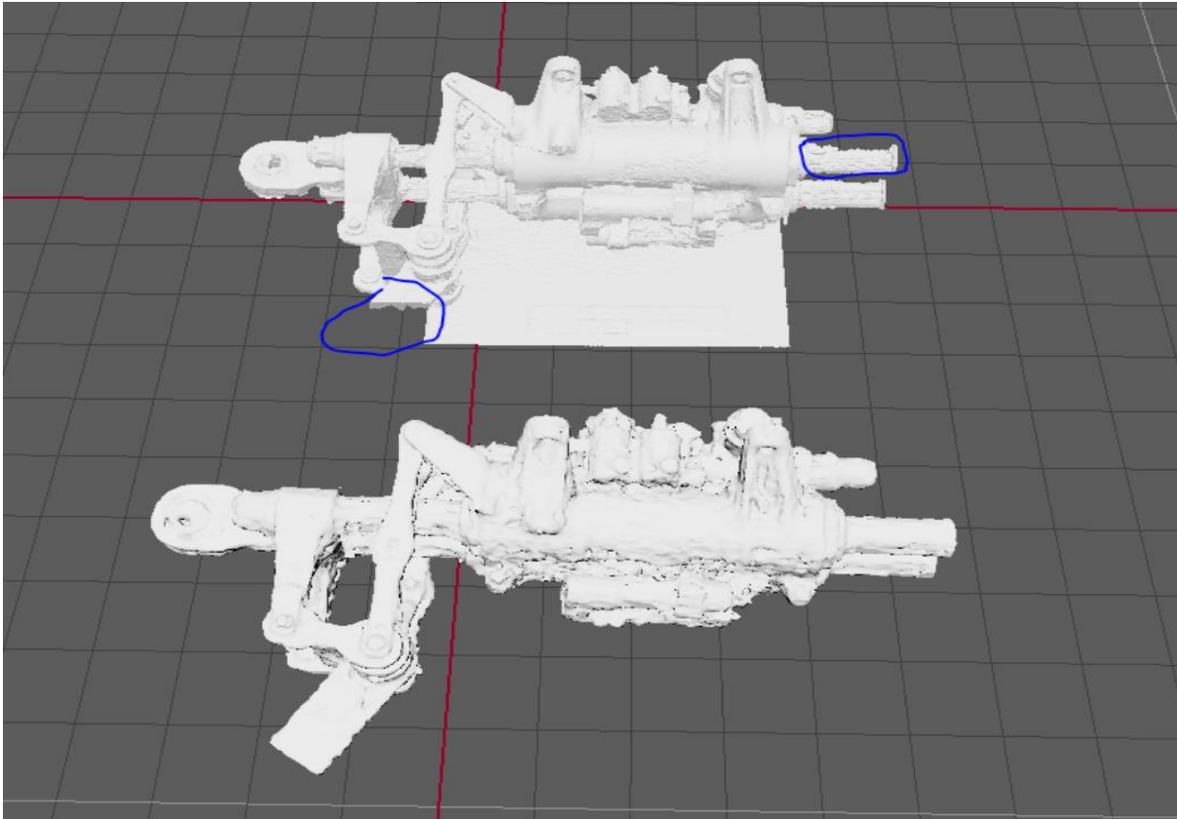
Figure 6: Mesh Models - NeRF Bottom, Photogrammetry Top

Figure 7: Rendered photogrammetry & NeRF Reference

How does mesh quality vary between photogrammetry and NeRF results on objects with only mild simple reflections? One would think that photogrammetry would handle more muted reflections better than NeRFs potentially, and in some cases it does. The manifolds of **Figure 7** were handled more precisely despite being a mix of dark and matte metal materials. However, the mildly reflective pipes still came out with a jagged inaccurate form. NeRF has achieved a much less noisy result on the reflective pipes. Textures and Materials tend to hide many mesh quality issues, so for the following example, we will show the results of the mesh exports of this object: Reality Capture and InstantNGP. As depicted in **Figure 8**, we can see what we have established with the prior asset example. Yes, the precision of the photogrammetry is generally higher, but even on mildly reflective pipes the photogrammetry created muddy results. This could still be used as a reference to model on top of as a 3D artist, but the NeRF shows that the method did an excellent job capturing the general form even in reflective areas such as on the pipes.
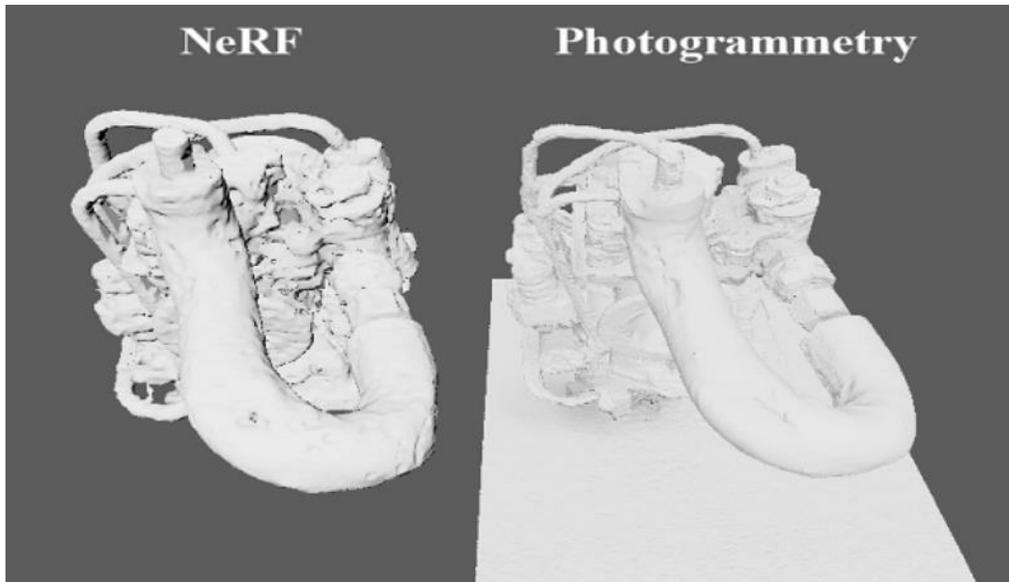
Figure 8: Simple Reflection Mesh Comparison

THE FUTURE OF NERF

It is important to note how quality NeRFs can be utilized beyond simply being a reference tool for 3D modelers and engineers. Several potential use cases for NeRFs go beyond using photogrammetry and NeRF together to fill in the gaps where each other fails or using NeRFs as a simple reference for 3D artists to model over and create a quality model capable of being rigged, animated, and textured. NeRFs and what comes from them will ultimately have more future applications than being used as a reference tool.

NeRF with 6 DoF Silhouette Recognition

Regarding 6 DoF silhouette recognition and tracking, a comprehensive form can sometimes be enough for a model target to work **(See Figure 9)**. Model targets represent an important technology behind recognizing objects via computer vision sensors and stabilizing and overlaying augmented reality projections to them. A model target can be described as a silhouette recognition profile created by loading CAD data or cleaning up SCAN data. The benefit of NeRF for capturing this model target that succeeds where photogrammetry fails is how well it captured the form even in highly metallic and reflective areas. It is worth mentioning that this model had some light cleanup done on it, but it was mostly just a vertex weld procedure and removing the table mesh from underneath it. Photogrammetry would have entirely omitted the metallic areas from the scan, which would have required significantly more model clean-up to recreate the reflective areas of the mesh. CAD will always be ideal over any form of scan data, but it is worth noting that CAD may not always be provided to a modeler, and this is why quality NeRF model targets can still be viable.
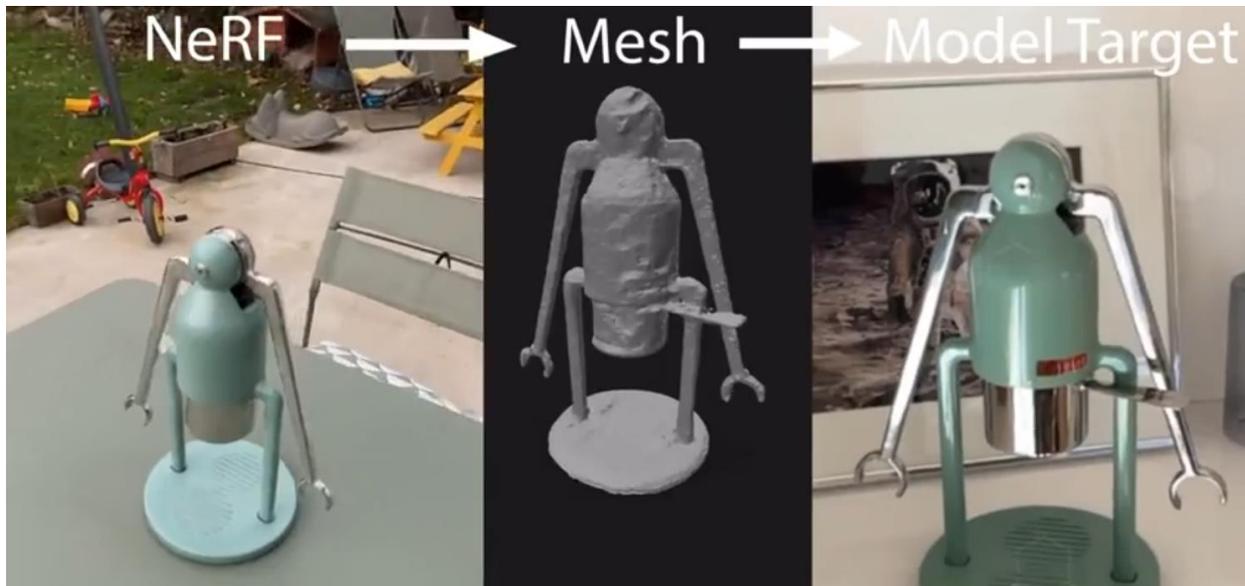
Figure 9: NeRF with Mixed Reality (Germain, 2022)

An interesting use case with NeRFs as they get more robust over time is the capability to remove objects digitally from the real world and place them elsewhere as seen in **Figure 10**. To do this, one needs a NeRF of the room and NeRFs of the objects moving around the world digitally. By digitally overlaying the NeRF environment over the real-world object, objects can appear to be digitally erased from their current position and moved elsewhere. There are numerous uses for NeRF applied in this way, such as interior planning for architects, loadmasters could use this to virtually move heavy components in a training scenario without the risk of having to move them physically, or maintenance workshops could use something like this for planning out different configurations before having to move everything around.

Figure 10: NeRF object replacement and removal. (Red shows what is under the virtual overlay) (Kari, Schütte, & Sodhi, 2023)

NeRFs with Virtual Reality

NeRFs can be utilized for object capture, mixed reality, pose estimation, and Virtual Reality/Environment Capture (**See Figure 11**). It is worth mentioning that similar alternatives are emerging that are more performant for Virtual Environments such as Gaussian Splats. Experimenting with additional capture techniques is a topic for another paper. With that in mind, as NeRFs and successor methods become more optimized and hardware improves, we will start

seeing the direct output of NeRF and Gaussian Splat environments directly integrated into game engines without a 3D modeler having to remodel everything.

Figure 11: VR environment capture (Instant NeRF VR First Impressions, 2023)

CONCLUSION

From a 3D artist's perspective, the topology of photogrammetry and NeRF models are not what we consider optimized, nor would the mesh exports be something that should be taken directly as is into a game engine. Game engines run well with optimized 3D mesh models that use enough polygons to accurately represent the subject's form without bloating the mode's polygon count. Neither photogrammetry nor NeRFs are perfect, and both ultimately export an entirely triangulated model that a 3D modeler could not texture, rig or animate. As this paper shows, NeRFs handle simple metallic reflections better, and when done correctly do capture the form of the object well. Photogrammetry does an excellent job of capturing the more matte details of an object's form with greater accuracy and precision. If neither method is perfect, yet both do something the other does not do well, it would be logical to use both to fill in the gaps where each other falls short. Artists can easily bring NeRFs and photogrammetry scans into their modeling applications and use them better to understand different areas of the object or environment. In the paper "Combining Neural Radiance Fields and Photogrammetry," the researchers showed that photogrammetry and NeRF can compensate for each other's weaknesses. (Croce, et al., 2024) The researchers went further by combining the two respective methods, but both can still be used together without combining them into one method. It may be easy to drift towards one or the other after looking at photogrammetry or NeRF, but if the data sets of photogrammetry can work well with NeRFs, then it makes more sense to support both to help create a more comprehensive result for 3D modelers.

NeRFs are also capable of being more than a simple reference tool. NeRFs can be used for 3D environments, Objects, Mixed Reality, Augmented Reality, and Virtual Reality. NeRFs and the new methods that come from them will inevitably improve. However, the current state of NeRFs is best as a reference tool used in tandem with photogrammetry until NeRF's pros and cons fully outweigh those of photogrammetry or a method better than either makes its way into the mesh and CAD modeling industries. Polygon limitations constitute a significant factor keeping photogrammetry and NeRFs from being performant in game engines. In addition, mesh topology from photogrammetry and NeRFs are not riggable for animation and do not unfold well for texturing. Until NeRF or photogrammetry form representations improve, we observed that using these methods as a reference for creating a 3D model rather than as a starting point is best.

REFERENCES

Croce, V., Billi, D., Caroti, G., Piemonte, A., De Luca, L., & Véron, P. (2024, January 11). *Comparative Assessment of Neural Radiance Fields and Photogrammetry in Digital Heritage: Impact of Varying Image Conditions on 3D Reconstruction.* Retrieved from MDPI: https://www.mdpi.com/2072-4292/16/2/301

di Filippo, A., Antinozzi, S., Cappetti, N., & Villecco, F. (2023). Methodologies for assessing the quality of 3D models obtained using close-range photogrammetry. *International Journal on Interactive Design and Manufacturing (IJIDeM)*.

Germain, F. (2022). *NeRF to model target — easy way to get shiny things into AR*. Retrieved from Reddit: https://www.reddit.com/r/AR_MR_XR/comments/ygj678/nerf_to_model_target_easy_way_to_get_shiny_things/

Instant NeRF VR First Impressions. (2023). Retrieved from YouTube: https://www.youtube.com/watch?v=u5x2D3Ejpnc&t=218s

Kari, M., Schütte, R., & Sodhi, R. (2023). Scene Responsiveness for Visuotactile Illusions in Mixed Reality. *ACM User Interface Software & Technology (UIST '23)*.

Mildenhall, B. S. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *In Proceedings of the European Conference on Computer Vision (ECCV)* , 405-421.

Müeller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, 102:1-102:15.

Pepe, M., Alfio, V., & Costantino, D. (2023). Assessment of 3D Model for Photogrammetric Purposes Using AI Tools Based on NeRF Algorithm. *Heritage*.

Remondino F, K. A. (2023, June 10). *A Critical Analysis of NeRF-Based 3D Reconstruction.* Retrieved from MDPI: https://www.mdpi.com/2072-4292/15/14/3585

Saif, W. (2022, November). *Photogrammetry: A Brief Historical Overview.* Retrieved from researchgate: https://www.researchgate.net/profile/Wahib-Saif/publication/365778003_Photogrammetry_A_Brief_Historical_Overview/links/6383aca9554def6193794e44/Photogrammetry-A-Brief-Historical-Overview.pdf

Yuan-Chen Guo, D. K.-H. (2021). *NeRFReN: Neural Radiance Fields with Reflections*. Retrieved from arxiv.org: https://arxiv.org/abs/2111.15234